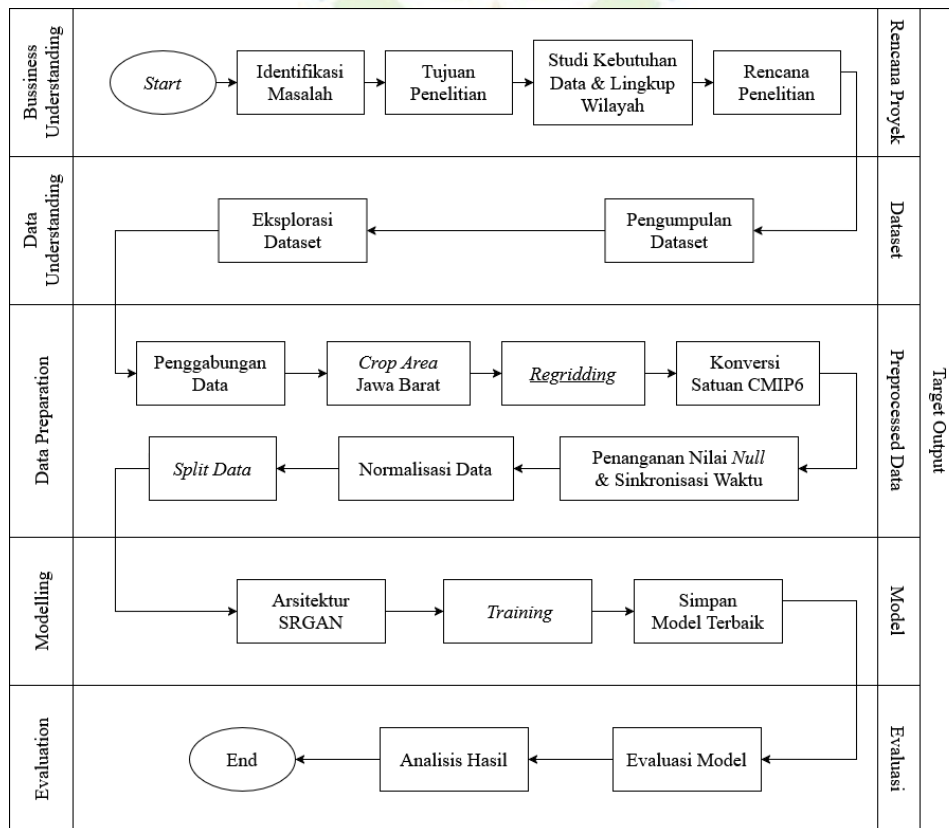


### BAB III METODOLOGI PENELITIAN

Bab ini memaparkan kerangka metodologis yang digunakan sebagai panduan dalam menyelesaikan permasalahan penelitian. Di dalamnya, diuraikan secara sistematis tahapan tahapan teknis dan alur kerja yang ditempuh untuk mencapai solusi yang optimal sesuai dengan tujuan yang telah ditetapkan. Penelitian ini mengadopsi kerangka kerja *Cross Industry Standard Process for Data Mining* (CRISP-DM) sebagai landasan metodologis yang bersifat iteratif. Pendekatan ini memberikan fleksibilitas untuk kembali ke tahap sebelumnya guna menyempurnakan model berdasarkan temuan pada tahap evaluasi. Alur kerja sistematis beserta rincian proses penelitian pada setiap fase disajikan pada Gambar 3.1.



Gambar 3.1 Alur Penelitian (CRISP-DM)

### 3.1 *Business Understanding*

#### 3.1.1 Identifikasi Masalah

Pada tahap ini akan dilakukan identifikasi permasalahan terkait penggunaan data *General Circulation Models* (GCM), khususnya *Coupled Model Intercomparison Project Phase 6* (CMIP6), untuk analisis curah hujan skala regional. Data GCM diketahui memiliki resolusi spasial yang rendah dan potensi bias terhadap data observasi.

Selain itu, akan dilakukan kajian terhadap pendekatan *downscaling* yang telah digunakan dalam penelitian terdahulu, termasuk metode statistik, *machine learning*, dan *deep learning*. Analisis ini akan digunakan untuk memahami celah penelitian yang masih terbuka dan menentukan pendekatan metodologis yang sesuai.

#### 3.1.2 Tujuan Penelitian

Berdasarkan permasalahan yang telah diidentifikasi, penelitian ini akan bertujuan untuk mengembangkan dan mengimplementasikan model berbasis SRGAN dalam melakukan *statistical downscaling* pada data curah hujan CMIP6.

Penelitian ini juga akan bertujuan untuk mengevaluasi kemampuan model dalam mentransformasikan data beresolusi rendah menjadi data beresolusi tinggi yang lebih representatif terhadap kondisi regional.

#### 3.1.3 Studi Kebutuhan Data

Pada tahap ini akan dilakukan penentuan wilayah studi dan kebutuhan dataset. Wilayah penelitian akan difokuskan pada Provinsi Jawa Barat. Dataset yang akan digunakan meliputi:

1. Data CMIP6 sebagai input beresolusi rendah.
2. Data *Climate Hazards Group InfraRed Precipitation with Station data Version 2* (CHIRPS2) sebagai data observasi (*ground truth*).
3. Data *Digital Elevation Model Nasional* (DEMNAS) sebagai variabel tambahan untuk merepresentasikan faktor topografi.

#### 3.1.4 Rencana Eksperimen

Bagian ini akan menyajikan rencana eksperimen yang menggambarkan alur penelitian mulai dari pengumpulan data hingga tahap evaluasi. Rencana

tersebut akan mencakup proses *data preparation*, strategi pelatihan model, serta evaluasi terhadap kinerja model, yang akan menjadi acuan dalam pelaksanaan penelitian pada Bab IV.

## **3.2 *Data Understanding***

### **3.2.1 Pengumpulan Dataset**

Pada tahap ini akan dilakukan pengumpulan data CMIP6, CHIRPS2, dan DEMNAS dari sumber resmi masing masing. Data akan disimpan dalam format yang sesuai, seperti NetCDF dan raster geospasial, serta disusun dalam struktur direktori yang terorganisir.

### **3.2.2 Eksplorasi Dataset**

Eksplorasi dataset akan dilakukan untuk memahami karakteristik awal data, yang meliputi:

1. Identifikasi resolusi spasial masing masing dataset
2. Analisis rentang waktu pengamatan
3. Pemeriksaan distribusi nilai curah hujan
4. Identifikasi nilai *null*
5. Visualisasi spasial awal

Hasil eksplorasi ini akan digunakan untuk menentukan strategi pemrosesan data pada tahap berikutnya.

## **3.3 *Data Preparation***

### **3.3.1 Penggabungan Data**

Penggabungan data akan dilakukan pada dataset DEMNAS dan CHIRPS. Data DEMNAS yang diperoleh tidak tersedia dalam satu file utuh untuk wilayah Jawa Barat, melainkan terbagi dalam beberapa tile sehingga perlu digabungkan menjadi satu file yang mencakup seluruh area studi.

Selain itu, data CHIRPS tersedia dalam format tahunan, sehingga diperlukan proses penggabungan antar tahun untuk membentuk satu dataset kontinu berdasarkan rentang waktu penelitian. Proses ini dilakukan agar data memiliki cakupan spasial dan temporal yang utuh sebelum digunakan pada tahap pemodelan.

### 3.3.2 *Crop Area Jawa Barat*

Seluruh dataset akan dipotong berdasarkan batas koordinat wilayah Jawa Barat guna membatasi area studi dan mengurangi beban komputasi.

### 3.3.3 *Regridding*

*Regridding* akan dilakukan untuk menyelaraskan resolusi spasial CMIP6 dengan grid CHIRPS2 agar diperoleh struktur grid yang konsisten.

### 3.3.4 *Konversi Satuan CMIP6*

Data curah hujan CMIP6 yang memiliki satuan berbeda akan dikonversi ke satuan yang seragam dengan CHIRPS2. Karena model hanya bisa membandingkan angka secara langsung, maka satuan harus diseragamkan terlebih dahulu agar proses *training* dan evaluasi menjadi valid.

### 3.3.5 *Penanganan Nilai Null dan Sinkronisasi Waktu*

Nilai *Null* akan diidentifikasi dan ditangani sesuai kebutuhan. Sinkronisasi waktu juga akan dilakukan untuk memastikan kesesuaian indeks waktu antara data CMIP6 dan CHIRPS2.

### 3.3.6 *Normalisasi*

Normalisasi dilakukan untuk menstabilkan distribusi nilai fitur agar proses pembelajaran model menjadi lebih optimal. Salah satu metode yang umum digunakan adalah dengan mengurangi setiap nilai dengan rata-rata, kemudian membaginya dengan standar deviasi data. Proses ini membuat skala antar fitur menjadi lebih konsisten, sehingga model dapat belajar dengan lebih efektif.

### 3.3.7 *Split Data*

Dataset yang telah diproses akan dibagi menjadi tiga bagian, yaitu *training set*, *validation set*, dan *test set*. Pembagian ini dilakukan berdasarkan rentang waktu data karena SRGAN membaca data iklim sebagai satu peta utuh wilayah Jawa Barat pada setiap harinya. Walaupun nanti tingkat kesalahannya dihitung pada setiap piksel, SRGAN pada dasarnya bekerja dengan melihat gambar wilayah secara keseluruhan.

### 3.4 *Modelling*

#### 3.4.1 **Arsitektur SRGAN**

Arsitektur SRGAN akan diimplementasikan menggunakan *framework deep learning*. Model akan terdiri atas dua komponen utama, yaitu *Generator* dan *Discriminator*. *Generator* akan dirancang untuk meningkatkan resolusi spasial data, sedangkan *Discriminator* akan dilatih untuk membedakan data hasil generasi dengan data observasi:

#### 3.4.2 **Training**

Model akan di *training* menggunakan kombinasi *loss function* dan *optimizer* yang sesuai. Parameter pelatihan yakni jumlah *epoch* dan *batch size* akan ditentukan berdasarkan kebutuhan eksperimen.

#### 3.4.3 **Simpan Model Terbaik**

Model dengan performa terbaik yakni dengan nilai *loss function* terendah akan disimpan untuk digunakan pada tahap evaluasi.

### 3.5 *Evaluation*

#### 3.5.1 **Evaluasi Model**

Model akan dievaluasi menggunakan beberapa metrik statistik, yaitu *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), dan *Pearson Correlation Coefficient* (PCC), untuk mengukur tingkat akurasi dan kesesuaian hasil *downscaling* terhadap data observasi.

#### 3.5.2 **Analisis Hasil**

Pada tahap ini performa model akan dianalisis berdasarkan metrik evaluasi statistik yang digunakan yakni:

1. RMSE untuk mengukur besarnya kesalahan prediksi secara keseluruhan.
2. MAE untuk mengukur rata-rata deviasi absolut antara hasil *downscaling* dan data observasi.
3. PCC untuk mengukur tingkat korelasi linier antara data hasil *downscaling* dan data observasi.

Hasil evaluasi akan digunakan untuk menilai tingkat akurasi dan kesesuaian hasil *downscaling* terhadap data observasi.